Econometrica, Vol. 46, No. 5 (September, 1978)

THE MEASUREMENT OF MOBILITY

BY A. F. SHORROCKS[1]

The paper explores some of the issues involved in constructing measures of mobility when the data are provided in the form of a transition matrix. An initial set of axioms is proposed which is inconsistent. They can, however, be reconciled if empirically unlikely transition matrices are eliminated from consideration. The paper then discusses the problem of comparing matrices not defined over the same interval. An index based on the convergence speed in a Markov chain process is able to compensate for differing time periods.

1. INTRODUCTION

IN DERIVING measures of inequality economists have been primarily interested in static distributions corresponding to a particular point in time. However, it is recognized that this does not provide the complete picture, since the relative positions of both individuals and firms are constantly changing. Thus our assessment of monopoly power in one particular industry is determined not only by the concentration of asset values or sales in a single year but also by the extent to which the same firms dominate the industry over time. Similarly, evidence on inequality of incomes or wealth cannot be satisfactorily evaluated without knowing, for example, how many of the less affluent will move up the distribution later in life.

These dynamic aspects of inequality have received little systematic attention. The economic literature which discusses mobility and makes some attempt at measurement broadly falls into two categories. In the first, elementary statistical techniques and indices such as the rank correlation coefficient are used to evaluate the changes in relative positions (Hart and Prais [**11**], Hymer and Pashigan [**13**], Joskow [**14**], Grossack [**8**], Singh and Whittington [**19**], Boyle and Sorenson [**6**], Whittington [**22**]). In the second category, measures of mobility are a by-product of simple stochastic specifications of changes over time (Adelman [**1**], Hart [**9, 10**]).

Many of these studies illustrate the dynamic movements with a matrix cross-classifying the states or classes occupied at two points in time. The same procedure is used for both incomes (Champernowne [**7**], Thatcher [**20**], Shorrocks [**18**]) and firms (Horowitz and Horowitz [**12**]), even when the measurement of mobility is not the primary aim. However, the largest number of applications of mobility tables occur in other social science disciplines where the states are taken to be social classes, geographical regions or occupational groups. It is here that the complementarity between static distributions and dynamic mobility is most fully appreciated and here, particularly with regard to intergenerational social mobility, that most of the proposed mobility indices are to be found.

[1] I should like to thank A. B. Atkinson, D. J. Bartholomew, J. Bibby, D. Jarratt, J. Psarris, M. D. Steuer, H. Theil, and two referees for helpful comments and suggestions.

The essential characteristics of this work may be adequately gauged from the recent contributions of Boudon [**5**], Bartholomew [**2**], and Bibby [**3**]. We make no attempt to duplicate this coverage, nor to describe the variety of alternative suggestions which have been advocated. However, it should be noted that the motivation behind these proposals seems to be dominated by one consideration: that they should have some intuitively appealing interpretation. There has been virtually no discussion of the properties exhibited by or desirable in such indices.[2]

The purpose of this paper is to explore some of the issues involved in the construction of mobility measures. In doing so we shall restrict ourselves to measures derived from a transition matrix P, the theoretical analogue of a mobility table whose rows have been appropriately scaled down to sum to unity. Thus p_{ij} represents the probability of transferring to state j for those starting in state i. Many of the proposed indices are directed at these matrices and they completely determine the dynamic structure if the process governing transitions follows a Markov chain.

We begin by presenting a number of properties which might be required of an index of mobility. Regarded as a set of axioms it is shown that they soon become inconsistent. There is a basic conflict between the assumption that the index should increase as the off-diagonal elements of P become larger and the notion of a perfectly mobile structure. The resolution of this conflict is discussed at some length and the remedy favored is to eliminate from consideration those transition matrices which will not arise in practice.

Section 3 is devoted to the problem of comparing structures whose transition matrices are defined over different time periods. This is of particular interest to economists since the interval between observations will frequently be dictated by the available data. Observed mobility depends in part on the innate mobility of the system and also on the length of the period during which potential changes can be converted into actual movements. Unless the contributions of these two factors can be disentangled many of the comparisons we may wish to make, for example between countries and over time, will be impossible. One measure which does allow adjustment for the influence of the time period is related to the speed of convergence in a Markov chain process towards its equilibrium distribution. This index resembles that suggested by Theil [**21**, Chapter 5] in the context of intergenerational social mobility.

2. PROPERTIES OF MOBILITY MEASURES

An index of mobility will be defined as a continuous real function $M(\cdot)$ over the set of transition matrices $\mathscr{P}$. We shall begin by restricting the range of the index to the interval [0, 1] as is the case with many static inequality measures.

(N) Normalization: $0 \leqslant M(P) \leqslant 1$ for all $P \in \mathscr{P}$.

[2] One exception is Bibby [**4**].

This imposes no significant constraint on the set of potential measures since a rank-preserving change of origin and scale can always be found such that the transformed function takes values within the chosen interval. The probability of movements between classes are given by the off-diagonal elements of the matrix. If one of these increases at the expense of the diagonal component we may regard the new structure as indicating a higher level of mobility and require the index to reflect this change accordingly. Writing $P > P'$ when $p_{ij} \geqslant p'_{ij}$ for all $i \neq j$ and $p_{ij} > p'_{ij}$ for some $i \neq j$, this can be expressed as

(M) Monotonicity: $P > P'$ implies $M(P) > M(P')$.

Acceptance of condition (M) immediately gives a quasi-ordering over the set $\mathcal{P}$, although by itself it is far too weak for practical purposes. It implies that the structure represented by the identity matrix will be ranked lower than any other transition matrix. Since the identity matrix arises when no transitions between strata take place at all, this is consistent with our a priori notions and we can associate the identity matrix with the minimum value of the index:

(I) Immobility: $\mathrm{M(I)} = 0$.

At the other end of the scale we can search for a matrix or matrices that might be supposed to exhibit maximum mobility. In this connection matrices with identical rows, so the probability of moving to any class is independent of that originally occupied, have been usually described as *perfectly mobile*. The use of this terminology, introduced by Prais [**16**], suggests that they should be assigned the maximum value of the index. Thus:

(PM) Perfect Mobility: $M(P) = 1$ if $P = \boldsymbol{u}\boldsymbol{x}'$ where $\boldsymbol{u} = (1, 1, \ldots, 1)'$ and $\boldsymbol{x}'\boldsymbol{u} = 1$.

Stricter versions of these last two properties are available. The index could be said to satisfy the *strong immobility* (SI) condition when $M(P) = 0$ if and only if $P = I$; and that of *strong perfect mobility* (SPM) when $M(P) = 1$ if and only if P has identical rows. These stronger versions rule out the possibility that other types of matrices can take the extreme values.

Whilst all of these properties have some appeal, none appears to command universal acceptance apart from (I), associating the completely immobile structure with a zero index value. As an illustration consider the measure proposed by Bartholomew [**2**, p. 24]:

$$M_B(P) = \sum_i p_i^* \sum_j p_{ij}|i - j|.$$

The summation over j gives the average number of class boundaries crossed by an individual originally in state i, and these are then weighted by the proportions in the corresponding equilibrium distribution $\boldsymbol{p}^*$, the solution of the equation $\boldsymbol{p}^* P = \boldsymbol{p}^*$. In general the measure satisfies (I) but not (SI), (N), (M), or (PM), although if the weights attached to the row summation were independent of the elements of P, both (SI) and (M) would also be valid.

It would be impossible to satisfy simultaneously all the above properties since (N), (M), and (PM) are incompatible. The proof is trivial. If we define

$$P_1=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \qquad P_2=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

then by (M), (PM), $M(P_2)>M(P_1)=1$ which violates (N). Assuming that a perfectly mobile structure is given the maximum value of the index, the precise range of the function is irrelevant. The basic conflict is thus between (PM) and (M).

This conflict can be resolved in a number of ways. A minor adjustment to (M), replacing $M(P)>M(P')$ with the weak inequality, will restore consistency at the expense of assigning the maximum mobility value to all members of the large class of matrices whose off-diagonal elements are no less than some perfectly mobile structure. Alternatively we may abandon (PM) and look for some other characterization of maximum mobility. If monotonicity (M) is retained, the index value one cannot be permitted for a matrix with a non-zero main diagonal. Consequently the matrix P_2 is the only candidate in a two state system. However, this structure could not be observed unless, during the corresponding time period, the numbers in the two categories are exactly reversed. This is clearly impossible if the original and final frequency distributions are roughly the same and significantly non-rectangular, so the mobility measure may be biased downwards in such cases. In addition, the argument for associating P_2 with maximum mobility fails to provide much guidance for an extension to three or more states. In relaxing (PM), therefore, we may lose sight of any objective notion of maximum mobility and have to rely instead on whatever one specific measure tells us is the most mobile structure.

There is another reason for being reluctant to abandon the perfect mobility condition. Interest in mobility is not only concerned with movement but also predictability—the extent to which future positions are dictated by the current place in the distribution. We can regard (M) and (PM) as distinguishing these separate aspects. For a completely immobile structure they are entirely in agreement: there is a total absence of movement and future positions are predetermined exactly. However, they diverge at the other extreme. When the measure increases with movement, as indicated by (M), matrices like P_2 are the most mobile. But P_2 represents a structure which is as rigid, in the sense of predictable, as the completely immobile case. If we concentrate instead on matrices which exhibit the least amount of predictability, then those which are perfectly mobile are the obvious choice. Thus from the viewpoint of predictability the restriction (PM) is not as artificial as it may appear at first sight.

That the movement and predictability issues are not always in accord only serves to illustrate further the clash between (M) and (PM). However, as more movement is observed it would be normal to expect the class occupied in the future to become less dependent on the present position. In general, therefore, they should be in harmony. This leads us to ask whether too much emphasis has been placed on examples of transition matrices which, by any stretch of the imagination, are unlikely to arise in practice. The matrix P_2 above is a case in point. If anything resembling such a structure were found for incomes or firm size it would be a remarkable discovery. Yet the argument for conflict between (M) and (PM) relies on that example or other unlikely structures. We shall therefore explore one final avenue for reconciling the two assumptions—that of restricting attention to transition matrices which stand a reasonable chance of being observed empirically.

In observed transition matrices the higher values tend to cluster about the main diagonal. The assumption that transition matrices have a dominant diagonal is, however, too strict. Slightly better is the requirement that the probability of remaining in the same category is no less than that of transferring to any other particular group. Formally $p_{ii} \geqslant p_{ij}$ for all i, j, when we shall say that P has a *maximal diagonal.* This can be weakened further to the condition that P has a *quasi-maximal diagonal* when there exists positive $\mu_1, \dots, \mu_n$ such that $\mu_i p_{ii} \geqslant \mu_j p_{ij}$ for all i, j. At this time we can only conjecture that all observed transition matrices have quasi-maximal diagonals, but inspection of examples given in the cited references (including social and occupational mobility tables) failed to reveal any violations.[3] Unfortunately, confirmation is occasionally tedious and it would be useful to have an equivalent definition which is easier to apply.[4]

Restricting the analysis to the subset $\mathscr{P}^*$ of $\mathscr{P}$ which has quasi-maximal diagonals has one important attraction: (M) and (PM) are no longer incompatible. For example the index[5]

$$\hat{M}(P) = \frac{n\text{-trace } P}{n-1},$$

where n is the number of states, clearly satisfies (I), (SI), and (M). Furthermore $\mathscr{P}^*$ contains all perfectly mobile matrices of the form $p_{ij} = q_j > 0$ (since we can choose $\mu_j = q_j^{-1}$) and for these trace $P = 1$, hence (PM). Finally

$$\mu_i p_{ii} \geqslant \mu_j p_{ij} \quad \text{for all } i, j$$

implies

$$p_{ii} \sum_j \mu_j^{-1} \geqslant \mu_i^{-1} \sum_j p_{ij} = \mu_i^{-1}$$

so

$$n \geqslant \text{trace } P = \sum_i p_{ii} \geqslant \frac{\sum_i \mu_i^{-1}}{\sum_j \mu_j^{-1}} = 1.$$

[3] For example, Theil [**21**, p. 246] reproduces a social mobility table in which 15 of the 100 elements violate the maximal diagonal condition. However, setting $\mu_1 = .25$, $\mu_2 = \mu_5 = \mu_6 = .5$, $\mu_{10} = 2.0$, and $\mu_j = 1$ otherwise, demonstrates that the matrix has a quasi-maximal diagonal.

[4] Separate necessary and sufficient conditions are shown in the Appendix.

[5] Prais [**16**] shows that the mean exit time from class i (or the average length of stay in class i) is given by $1/(1-p_{ii})$. Since $\hat{M}$ can be rewritten as $\hat{M}(P) = \Sigma_i (1-p_{ii})/(n-1)$ it is the reciprocal of the harmonic mean of the mean exit times, normalized by the factor $n/(n-1)$.

From this (N) follows and also (SPM), since we have the equality $p_{ii}\,\Sigma_j\,\mu_j^{-1}=\mu_i^{-1}$ when $\hat{M}(P)=1$. Then

$$p_{ij}\leqslant\frac{\mu_j^{-1}p_{ii}}{\mu_i^{-1}}=\frac{\mu_j^{-1}}{\sum_j\mu_j^{-1}}\qquad\text{for all } i, j$$

and, as the rows of P sum to unity, $p_{ij}=\mu_j^{-1}/\Sigma_j\,\mu_j^{-1}$. Thus for $P\in\mathcal{P}^*$, $\hat{M}(P)$ satisfies all the properties mentioned earlier. Naturally this does not exclude the possibility that other measures have all these properties.

3. THE OBSERVATION PERIOD

We now turn to the time period over which the transition matrix is defined. The question is whether we can hope to make mobility comparisons if the observation periods corresponding to the matrices are not identical. Unless some adjustment is made for different time intervals, there will be a tendency to give an inflated mobility value to the structure defined over the longer period. In a short space of time there is little opportunity for movement, even if the structure is inherently very mobile.

The need to disentangle the effect of time does not appear to have been appreciated. This is due primarily to the focusing of attention on social mobility, where the accepted time interval is one generation. In other applications, especially those of interest to economists, there is no correspondingly obvious choice. We may therefore be faced with, say, ranking the mobility of firms in the U.S. and U.K. when one set of data is a cross-tabulation of asset values in consecutive years and the other a cross-tabulation for years separated by a longer period.

As an illustration of the difficulties which can arise, consider the transition matrices

$$P^1=\begin{bmatrix}.9&.1&.0\\.3&.4&.3\\.3&.3&.4\end{bmatrix},\qquad Q^2=\begin{bmatrix}.44&.28&.28\\.28&.44&.28\\.28&.28&.44\end{bmatrix},$$

where P^1 refers to a one year interval and Q^2 to two years. A comparison using $\hat{M}$ reveals that $\hat{M}(P^1)=.65<.84=\hat{M}(Q^2)$. However, we expect to observe more mobility over a longer period, so we are not justified in concluding that the structure which generated Q^2 is more mobile than that underlying P^1. To make a valid comparison additional assumptions must be introduced, and the natural approach would be to suppose that the transitions indicated by P^1 are repeated for a further year.[6] This allows us to obtain the corresponding two year transition matrix

$$P^2=P^1P^1=\begin{bmatrix}.84&.13&.03\\.48&.28&.24\\.48&.27&.25\end{bmatrix}.$$

[6] In effect this assumes that the process is a Markov chain with transition matrix P^1.

Now we have standardized for the different time period and $\hat{M}(P^2) = .815 < \hat{M}(Q^2)$ suggests that the system generating Q^2 is the more mobile.

Unfortunately, that is not the end of the matter. It could equally well have been argued that Q^2 was a two year transition matrix for a Markov chain process and that we should look for the associated one year matrix Q^1 to compare with P^1. Such a matrix is given by

$$Q^1 = \begin{bmatrix} .6 & .2 & .2 \\ .2 & .6 & .2 \\ .2 & .2 & .6 \end{bmatrix}$$

and $\hat{M}(Q^1) = .6 < \hat{M}(P^1)$, reversing the rankings previously assigned to the systems generating the matrices under examination. We are thus unable to say whether P^1 or Q^2 represents the more mobile structure.

The inability to make unambiguous rankings from transition matrices defined over different time periods is common to almost all the mobility measures ever proposed. This can be demonstrated with relatively simple examples. However, the problem does not arise if the index satisfies:

(PC) Period Consistency: $M(P) \geqslant M(Q)$ implies $M(P^k) \geqslant M(Q^k)$ for all integers $k \geqslant 1$.

This is certainly a desirable property. Without it we may be forced to conclude that mobility comparisons are impossible unless data are available for the same time period. Yet it places a severe restriction on the class of potential measures and eliminates the vast majority of those in current use. Moreover we once again run into a conflict with the monotonicity condition (M), since the matrix P_2 given earlier becomes $P_2^2 = I$ over two periods and hence $M(P_2) = M(I)$ by (PC), violating (M) and (SI). As was argued in the last section, this particular difficulty might be overcome by restricting the permissible domain of transition matrices under consideration.

It is not easy to understand the nature of the restriction which period consistency imposes on mobility measures. Nor does it achieve the original objective of separating the inherent mobility of a structure from the influence of time. We shall therefore explicitly introduce the time period T, over which the transition matrix P is defined, into the index to give $M(P; T)$ and replace (PC) with the stronger condition:

(PI) Period Invariance: $M(P; T) = M(P^k; kT), \quad k \geqslant 1.$

The advantage of this formulation is that the index now compensates for the length of the time interval. If the process is a Markov chain, the mobility value obtained for any structure will be independent of the particular observation period prescribed by the data.

Period invariance requires an index to combine, either explicitly or implicitly, the characteristic roots of the transition matrix in a specific way. Denote the roots by λ_i ($i=1,\ldots,n$) and order them so that $1=\lambda_1 \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_n|$. Applying the standard decomposition[7]

$$P=\sum_i \lambda_i E_i,$$

where

$$E_i E_j = \begin{cases} 0, & i \neq j, \\ E_i, & i=j, \end{cases}$$

and

$$\sum_i E_i = I,$$

we obtain

$$P^k=\sum_i \lambda_i^k E_i.$$

Now write $M(P;T)$ as the equivalent function $f(T,\lambda_2,\ldots,\lambda_n, E_1,\ldots,E_n)$. Then

$$\begin{aligned} f(T,\lambda_2,\ldots,\lambda_n,E_1,\ldots,E_n) &= M(P;T) \\ &= M(P^k;kT) \\ &= f(kT,\lambda_2^k,\ldots,\lambda_n^k,E_1,\ldots,E_n). \end{aligned}$$

The index must be an even function of each eigenvalue argument and is homogeneous of degree zero in the time period and logarithms of the characteristic roots. Notice that the properties considered in Section 2, with the exception of monotonicity, can also be expressed conveniently in terms of the characteristic roots. Suppressing the spectral matrices E_i, we have

(PI) $\quad f(T,\lambda_2,\ldots,\lambda_n)=f(kT,\lambda_2^k,\ldots,\lambda_n^k)$,

(I) $\quad f(T,1,1,\ldots,1)=0$,

(PM) $\quad f(T,0,0,\ldots,0)=1$,

and the stronger versions

(SI) $\quad f(T,\lambda_2,\ldots,\lambda_n)=0 \quad$ iff $\quad \lambda_2=\ldots=\lambda_n=1$,

(SPM) $\quad f(T,\lambda_2,\ldots,\lambda_n)=1 \quad$ iff $\quad \lambda_2=\ldots=\lambda_n=0$.[8]

[7] See, for example, Perlis [**15**, p. 173]. The matrix E_i is formed by the product of the right column eigenvector $\boldsymbol{x}_i$ and left row eigenvector $\boldsymbol{y}_i'$ corresponding to λ_i, under the normalization $\boldsymbol{y}_i'\boldsymbol{x}_i=1$. We shall discount the possibility that such a decomposition is not available since this can only occur when P has repeated roots. Under the usual topology, matrices with distinct roots form a dense subset within the set of all matrices.

[8] Since the roots have been ordered, this is equivalent to $\lambda_2=0$.

There are two indices satisfying (PI) which are closely related to measures discussed in the literature. The first is

$$M_D(P;\, T) = 1 - |\det P|^{\alpha/T}, \qquad \alpha > 0.$$

Measures derived from the determinant have not been well received, on the basis that they give the completely mobile value when any two rows (or columns) of the matrix are identical. Returning to our earlier discussion, however, we might question whether these hypothetical cases are sufficient grounds for dismissal. Moreover M_D has a further interesting property for non-homogeneous systems. If we assume that the process is a Markov chain with transition matrices P for the period $[0, T]$ and Q for the interval $[T, 2T]$, then the transition matrix over $[0, 2T]$ is PQ. Hence

$$\begin{aligned} 1 - M_D(PQ;\, 2T) &= |\det PQ|^{\alpha/2T} \\ &= [|\det P|^{\alpha/T} |\det Q|^{\alpha/T}]^{\frac{1}{2}} \\ &= [(1 - M_D(P;\, T))(1 - M_D(Q;\, T))]^{\frac{1}{2}}. \end{aligned}$$

Thus $1 - M_D$ (which can be regarded as an index of *immobility* or *rigidity*) for the combined period is the geometric mean of the subperiod rigidity values. When the dynamic structure can be appropriately specified as a Markov chain, mobility over extended lengths of time may be deduced from knowledge only of the index values for shorter intervals.

A second index having the property of period invariance is

$$M_H(P;\, T) = e^{-hT}$$

where

$$h = \frac{-\log 2}{\log |\lambda_2|}.$$

The expression given by h indicates the speed of convergence towards the equilibrium distribution for a Markov chain with transition matrix P. The close correspondence between mobility and the convergence speed has been noted both in Theil [**21**, Chapter 5] and Shorrocks [**17**, Section 2.2]. Intuitively a rigid structure is associated with a slowly changing distribution and convergence is comparatively slow. On the other hand a perfectly mobile structure establishes the equilibrium distribution within a single period. These give the extreme values of H as 0 (when $\lambda_2 = 0$) and ∞ (when $|\lambda_2| = 1$), so $M_H(P)$ lies in the interval $[0, 1]$.

The precise interpretation of h is the *asymptotic half life* defined as follows. Suppose that $p_i(t)$ represents the proportion of individuals in class i after t periods, each of length T years, and that $\boldsymbol{p}(t) = (p_1(t), \ldots, p_n(t))$ is generated by the homogeneous Markov chain

$$\begin{aligned} \boldsymbol{p}(t) &= \boldsymbol{p}(t-1)P \\ &= \boldsymbol{p}(0)P^t. \end{aligned}$$

For simplicity, suppose also that $1 = \lambda_1 > |\lambda_2| > |\lambda_3|$. Then the system has a unique stationary state distribution $\boldsymbol{p}^*$ and deviations from equilibrium follow

$$\begin{aligned}\boldsymbol{d}(t) &= \boldsymbol{p}(t) - \boldsymbol{p}^* \\ &= \boldsymbol{p}(0)P^t - \boldsymbol{p}^* \\ &= \sum_{i=1}^{n} \lambda_i^t \boldsymbol{p}(0)E_i - p^* \\ &= \sum_{i=2}^{n} \lambda_i^t \boldsymbol{p}(0)E_i\end{aligned}$$

since the first spectral matrix $E_1 = \boldsymbol{up}^*$.

The time taken to converge to within half the distance from equilibrium, commencing from the distribution $\boldsymbol{p}(t)$, is the minimum value of h such that

$$\frac{\|\boldsymbol{d}(t+h)\|}{\|\boldsymbol{d}(t)\|} \leqslant \frac{1}{2}$$

where $\|\cdot\|$ indicates the chosen metric. In general the "half life" h will depend both on the metric and the initial distribution,[9] but if we restrict ourselves to a linear metric (a norm) there is a unique asymptotic half life obtained by letting $t \to \infty$. Since λ_2 dominates the remaining eigenvalues

$$\boldsymbol{d}(t) \to \lambda_2^t \boldsymbol{p}(0)E_2$$

and

$$\frac{\|\boldsymbol{d}(t+h)\|}{\|\boldsymbol{d}(t)\|} \to |\lambda_2|^h.$$

Setting this equal to $\frac{1}{2}$ gives the expression for h. Converting the units, this becomes a half life of hT years.

The half life measure M_H is period invariant since

$$\begin{aligned}M_H(P^k; kT) &= \exp\left\{-\frac{kT \log 2}{\log |\lambda_2^k|}\right\} \\ &= \exp\left\{-\frac{T \log 2}{\log |\lambda_2|}\right\} = M_H(P; T).\end{aligned}$$

Clearly it also satisfies (N), (I), (PM), and (SPM). However it will, in general, violate both (SI) and (M). It seems unlikely that the latter difficulty can be overcome by an acceptable restriction on $\mathscr{P}$. Moreover, accurate computation of eigenvalues requires sophisticated programs not always available and very little is known about the sensitivity to sampling errors. For M_H, therefore, the attractiveness of the period invariance property is offset by other disadvantages. Other

[9] Theil [**21**, pp. 261–265] uses a quadratic approximation to his "informational distance measure" as the chosen metric, when convergence behavior is examined. He proposes λ_2^2 as a "measure for the imperfection of intergenerational social mobility" (p. 265).

period invariant measures remain to be investigated, but we may finally have to admit that no single mobility statistic has the minimum requirements regarded as essential.

London School of Economics

Manuscript received March 1976; final revision received July, 1977.

APPENDIX

An $n \times n$ non-negative matrix P has a quasi-maximal diagonal if there exists $\mu_1, \ldots, \mu_n > 0$ such that

$$\mu_i p_{ii} \geqslant \mu_j p_{ij} \qquad \text{for all } i, j.$$

THEOREM 1: *If a transition matrix P has a q.m.d. then* (a) $p_{ii} > 0$ *for all* i; (b) *trace* $P \geqslant 1$; (c) *all second order principal minors are non-negative.*

PROOF: (a) $p_{ii} = 0$ implies $0 = \mu_i p_{ii} \geqslant \mu_j p_{ij}$ for all j, hence $p_{ij} = 0$ for all j and this contradicts $\Sigma_j p_{ij} = 1$.
(b) Proof given in text.
(c) For all i, j, $\mu_i p_{ii} \geqslant \mu_j p_{ij}$ and $\mu_j p_{jj} \geqslant \mu_i p_{ji}$. Hence

$$\frac{p_{ii}}{p_{ij}} \geqslant \frac{\mu_j}{\mu_i} \geqslant \frac{p_{ji}}{p_{jj}}$$

and

$$\begin{vmatrix} p_{ii} & p_{ij} \\ p_{ji} & p_{jj} \end{vmatrix} \geqslant 0.$$

THEOREM 2.: *If P is a transition matrix with $p_{ii} > 0$ for all i, then sufficient conditions for q.m.d. are* (a) $p_{ii} \geqslant p_{ji}$ *for all* i, j, *or* (b) *for all* $i > 1$, $p_{ij} > 0$ *for some* $j < i$ *and*

$$\begin{vmatrix} p_{ik} & p_{im} \\ p_{mk} & p_{mm} \end{vmatrix} \geqslant 0 \qquad \textit{for all } i,\ k < m \leqslant n.$$

PROOF: (a) Choose $\mu_i = 1/p_{ii} > 0$. Then

$$\mu_i p_{ii} = \mu_j p_{jj} \geqslant \mu_j p_{ij} \qquad \text{for all } i, j.$$

(b) Choose $\mu_1 = 1$ and define recursively $\mu_i = \max_{i>j} \{\mu_j p_{ij}/p_{ii}\} > 0$. Then $\mu_i p_{ii} \geqslant \mu_j p_{ij}$ for all $i \geqslant j$ and for every $i > 1$ equality holds for at least one $j < i$. Now $\mu_i p_{ii} \geqslant \mu_1 p_{i1}$ for all i, so suppose that for all $j \leqslant m-1$, $\mu_i p_{ii} \geqslant \mu_j p_{ij}$ for all i. For some $k < m$, $\mu_m p_{mm} = \mu_k p_{mk}$. Thus, for all i,

$$\mu_i p_{ii} \geqslant \mu_k p_{ik}$$

and, when $m > i$,

$$\mu_i p_{ii} \geqslant \mu_k (p_{mk} p_{im}/p_{mm}) \qquad \text{since} \qquad \begin{vmatrix} p_{ik} & p_{im} \\ p_{mk} & p_{mm} \end{vmatrix} \geqslant 0,$$

$$= \mu_m p_{im}.$$

But given the choice of μ_i, $\mu_i p_{ii} \geqslant \mu_m p_{im}$ for all $i \geqslant m$. So $\mu_i p_{ii} \geqslant \mu_m p_{im}$ for all i and, by induction on m, P has a q.m.d.

NOTE: The sufficiency condition (a) is easy to confirm by inspection and holds for a large number of transition matrices. If the diagonal components are positive and the probability of remaining in class i is at least as great as that of moving into the class from any other category, then the matrix has a quasi-maximal diagonal. That neither (a) nor (b) is necessary can be seen from the following example with a maximal diagonal:

$$\begin{bmatrix} .6 & .4 & .0 \\ .0 & .5 & .5 \\ .3 & .3 & .4 \end{bmatrix}.$$

REFERENCES

[1] ADELMAN, I. G.: "A Stochastic Analysis of the Size Distribution of Firms," *Journal of the American Statistical Association*, 53 (1958), 893–904.
[2] BARTHOLOMEW, D. J.: *Stochastic Models for Social Processes*, Second Edition. London: Wiley, 1973.
[3] BIBBY, J.: "Methods of Measuring Mobility," *Quality and Quantity*, 9 (1975), 107–136.
[4] ———: "Some Recent Work on Social Mobility Tables," paper presented to the SSRC Seminar on Occupational Mobility, September, 1975.
[5] BOUDON, R.: *Mathematical Structures of Social Mobility*. Amsterdam: Elsevier, 1973.
[6] BOYLE, S. E., AND R. L. SORENSEN: "Concentration and Mobility: Alternative Measures of Industry Structure," *Journal of Industrial Economics*, 19 (1971), 118–132.
[7] CHAMPERNOWNE, D. G.: "A Model of Income Distribution," *Economic Journal*, 63 (1953), 318–351.
[8] GROSSACK, I. M.: "Towards an Integration of Static and Dynamic Measures of Industry Concentration," *Review of Economics and Statistics*, 47 (1965), 301–308.
[9] HART, P. E.: "Comparative Statics and Dynamics of Income Distributions," *Journal of the Royal Statistical Society*, Series A, Part 1, 139 (1976), 108–125.
[10] ———: "The Dynamics of Earnings, 1963–1973," *Economic Journal*, 86 (1976), 551–565.
[11] HART, P. E., AND S. J. PRAIS: "The Analysis of Business Concentration: A Statistical Approach," *Journal of the Royal Statistical Society*, Series A, Part II, 119 (1956), 150–191.
[12] HOROWITZ, A., AND I. HOROWITZ: "Entropy, Markov Processes and Competition in the Brewing Industry," *Journal of Industrial Economics*, 16 (1968), 196–211.
[13] HYMER, S., AND P. PASHIGAN: "Turnover of Firms as a Measure of Market Behavior," *Review of Economics and Statistics*, 44 (1962), 82–87.
[14] JOSKOW, J.: "Structural Indicia: Rank-Shift Analysis as a Supplement to Concentration Ratios," *Review of Economics and Statistics*, 42 (1960), 113–116.
[15] PERLIS, S.: *Theory of Matrices*, Reading, Mass: Addison-Wesley, 1952.
[16] PRAIS, S. J.: "Measuring Social Mobility," *Journal of the Royal Statistical Society*, Series A, Part I, 118 (1955), 56–66.
[17] SHORROCKS, A. F.: "Aspects of the Distribution of Personal Wealth," Ph.D. Thesis, London University, 1973.
[18] ———: "Income Mobility and the Markov Assumption," *Economic Journal*, 86 (1976), 566–578.
[19] SINGH, A., AND G. WHITTINGTON: *Growth, Profitability and Valuation*. Combridge: Cambridge University Press, 1968.
[20] THATCHER, A. R.: "Year-to-year Variations in the Earnings of Individuals," *Journal of the Royal Statistical Society*, Series A, Part III, 134 (1971), 374–382.
[21] THEIL, H.: *Statistical Decomposition Analysis*. Amsterdam: North Holland, 1972.
[22] WHITTINGTON, G.: "Changes in the Top 100 Quoted Manufacturing Companies in the United Kingdom 1948 to 1968," *Journal of Industrial Economics*, 21 (1972), 17–34.